

一种改进的基于翻译的知识图谱表示方法

方阳¹ 赵翔^{1,2} 谭真¹ 杨世宇³ 肖卫东^{1,2}

¹(国防科技大学信息系统与管理学院 长沙 410073)

²(地球空间信息技术协同创新中心(武汉大学) 武汉 430079)

³(新南威尔士大学计算机科学与工程学院 澳大利亚悉尼 2052)

(fangyang12@nudt.edu.cn)

A Revised Translation-Based Method for Knowledge Graph Representation

Fang Yang¹, Zhao Xiang^{1,2}, Tan Zhen¹, Yang Shiyu³, and Xiao Weidong^{1,2}

¹(College of Information System and Management, National University of Defense Technology, Changsha 410073)

²(Collaborative Innovation Center of Geospatial Technology (Wuhan University), Wuhan 430079)

³(School of Computer Science and Engineering, the University of New South Wales, Sydney, Australia 2052)

Abstract Knowledge graph is of great research value to artificial intelligence, which has been extensively applied in the fields of semantic search and question answering, etc. Knowledge graph representation transforms a large-scale knowledge graph comprising entities and relations into a continuous vector space. To this end, there have been a number of models and methods proposed for knowledge embedding. Among them, TransE is a classic translation-based method that is of low model complexity, high computational efficiency, as well as good capability of expressing knowledge. However, TransE still has two flaws: one is that it utilizes inflexible Euclidean distance as metric, and treats each feature dimension identically, hence, the model accuracy may be interfered by irrelevant dimensions; the other is that it has limitations in dealing with complex relations including reflexive, one-to-many, many-to-one and many-to-many relations. Currently, there has not been a single method that resolves the flaws simultaneously, and thus, we propose a revised translation-based method for knowledge graph representation, namely, TransAH. For the first flaw, TransAH adopts an adaptive metric, replacing Euclidean distance with weighted Euclidean distance by adding a diagonal weight matrix, which assigns different weights to every feature dimension. As to the second, inspired by TransH, it introduces the relation-oriented hyperspace model, projecting head and tail entities to hyperspace of a given relation for distinction. At last, empirical studies on public real knowledge graph datasets analyze and verify the effectiveness of the proposed method. Comprehensive comparative experiments using two tasks-link prediction and triplet classification show that, in contrast to the existing models and methods, TransAH achieves remarkable improvement in various aspects and demonstrates its superiority.

Key words knowledge graph; knowledge representation; representation learning; link prediction; triplet classification

收稿日期:2016-09-27 修回日期:2017-02-21

基金项目:国家自然科学基金项目(61402494,61402498,71690233);湖南省自然科学基金项目(2015JJ4009)

This work was supported by the National Natural Science Foundation of China (61402494, 61402498, 71690233) and the Natural Science Foundation of Hunan Province (2015JJ4009).

通信作者:赵翔(xiangzhao@nudt.edu.cn)

摘要 知识图谱在人工智能上有很大的研究价值,并被广泛应用于语义搜索和自动问答等领域.知识图谱表示将包含了实体和关系的大规模知识图谱映射到一个连续的向量空间.为此,有一系列知识表示模型提出,其中基于翻译模型的经典方法 TransE 不仅模型复杂度低、计算效率高,而且同样具有良好的知识表达能力.但是,TransE 亦存在 2 个缺陷:1)它使用了不够灵活的欧氏距离作为度量,对每一个特征维同等对待,模型的准确性可能受到无关维度的干扰;2)它在处理自反、一对多、多对一和多对多等复杂关系时存在局限性.目前,还没有一种方法能同时解决上述 2 个缺陷,因此提出一种改进的基于翻译的知识图谱表示方法 TransAH.对于第 1 个缺陷,TransAH 采用了一种自适应的度量方法,加入了对角权重矩阵将得分函数中的度量由欧氏距离转换为加权欧氏距离,并实现了为每一个特征维区别地赋予权重.针对第 2 个缺陷,受 TransH 方法的启发,TransAH 引入面向特定关系的超平面模型,将头实体和尾实体映射至给定关系的超平面加以区分.最后,在公开真实的知识图谱数据集上分析和验证了所提方法的有效性.利用链路预测和三元组分类这 2 项任务开展了全面横向评测实验,相较于现有的模型和方法,TransAH 在各项指标上均取得了很大的进步,体现了其优越性.

关键词 知识图谱;知识表示;表示学习;链路预测;三元组分类

中图法分类号 TP391

知识图谱是结构化的语义知识库,以符号形式描述物理世界中的实体(entities)及其相互关系(relations),其基本组成单元是“实体-关系-实体”三元组(triplets)以及实体和相关属性的值对;实体间通过关系相互联结,构成网状的知识结构.

诸如 Freebase^[1] 和 WordNet^[2] 等大型知识图谱在人工智能领域方面呈现出广泛的应用价值,被用于支撑语义检索以及自动问答等高级应用^[3].例如在检索信息时,用户的查询词是典型的短文本,一个查询词往往仅由几个关键词构成.传统的关键词匹配检索技术不理解查询词背后的语义信息,查询结果可能会很差.为此,人们一直在探索比关键词查询更高效的信息检索方式,而构建大规模知识图谱使得语义检索成为可能.语义检索能够更好地理解用户的查询词,从纷繁复杂的信息中有效筛选出那些最合适的答案,进而达到关键词匹配检索无法匹敌的效果.又如知识图谱可作为自动问答系统的知识库,通过知识图谱中实体的复杂关系推理得到问题的答案.无论是理解用户查询意图,还是自动寻求问题答案,都毫无例外地需要进行语义理解和知识推理,而这些智能技术取得巨大进展的背后则是更深、更广、更新和更加准确的知识图谱的构建和运用.

知识图谱领域的主要研究目标是从无(半)结构的互联网信息中获取有结构知识,自动融合构建知识库、服务知识推理等相关应用.其中,知识表示是知识获取与应用的基础,是贯穿知识库构建与应用

全过程的关键问题.表示知识图谱最直接的方法是利用图数据库^[4],但是这种表示手段应用在大规模知识图谱上存在计算复杂度高、推理效率低和数据稀疏等问题.换句话说,在这种表示之下,知识图谱是符号化的,并具备有逻辑性,因此数值化的机器学习方法和技术均不能应用到知识图谱上.

近年来,随着大数据研究与应用的不断深入,人工智能中的表示学习技术异军突起,旨在将研究对象的语义信息表示为稠密低维实值向量.面向知识图谱的表示学习作为一种支持知识图谱计算和推理的新方法,在保留原始图谱特定属性的同时,将知识图谱映射为连续的向量空间,使得一大批高效的数值化计算和推理方法得以适用.因此,知识图谱的低维向量表示的是一种分布式表示(distributed representation),即孤立地看表示向量中的每一维,它表达了一种没有明确对应含义的潜在特征(称作“特征维”);但综合各维形成一个向量,则能够表示对象的语义信息.

鉴于上述优点,研究者提出了若干知识图谱表示模型,包括基于翻译(translation-based)的模型、结构嵌入(structured embedding, SE)模型、语义匹配能量(semantic matching energy, SME)模型和潜变量模型(latent factor model, LFM)等.本文主要考虑基于翻译模型的知识图谱表示方法.

具体地,知识图谱中的一条知识通常由三元组(h, r, t)表示,其中 h 是头实体, t 是尾实体, r 表示头实体和尾实体之间的关系.知识图谱表示用一个

k 维向量 \mathbf{h}^* (或者 \mathbf{t}^*) 来代表头(或者尾)实体, 用一个转换向量 \mathbf{r}^* 来表示头尾实体对之间的关系^①; 同时通过定义一个得分函数 $f_r^*(\mathbf{h}^*, \mathbf{t}^*)$ 来衡量三元组 (h, r, t) 在表示空间中成立的可能性. 实体和关系的表示模型则是通过最小化包含所有实体和关系的全局得分函数实现的. 因此, 即使是一个单一的实体或关系的表示都可以捕捉到整个知识图谱的全局信息.

现有方法中, TransE 是一个具有代表性的经典方法^[5], 其得分函数是 $f_r^*(\mathbf{h}^*, \mathbf{t}^*) = \|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*\|^2$, 即在表示空间中头实体经过关系转换之后与尾实体之间的欧氏距离. TransE 在取得较好预测表现的同时保持了足够的简洁性和高效性, 但存在至少 2 个方面的缺陷: 缺陷 1 是 TransE 使用欧氏距离作为得分函数中的度量, 每一个特征维以相同的权重参与计算, 有失灵活, 知识表示的准确性亦可能受到无关维度的影响; 缺陷 2 是它在处理自反、一对多、多对一以及多对多^②等复杂关系时存在局限性, 不能良好区分具有相同关系的实体. 然而, 基于笔者目前所知, 暂无一种知识图谱表示方法同时解决了上述 2 个方面的缺陷.

为此, 本文提出了一种改进的知识图谱表示方法 TransAH, 同时针对上述 2 方面缺陷提升模型表示知识的能力. 具体地, 针对第 1 个缺陷, 提出在得分函数中添加一个对角权重矩阵 \mathbf{D}_r^* , 即 $(\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*)^T \mathbf{D}_r^* (\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*)$, 将欧氏距离的度量转换为更为灵活的加权欧氏距离, 实现对不同的特征维度赋予自适应的权重. 采用对角权重矩阵, 不仅避免了无关特征维度的干扰, 同时启发式地大幅提高了模型训练时的计算效率. 其次, 针对第 2 个缺陷, 受 TransH 方法启发, 采用了面向关系的超平面模型, 将三元组中的 \mathbf{h}^* 和 \mathbf{t}^* 映射到给定关系的超平面上以有效表征复杂关系. 将 \mathbf{h}^* 和 \mathbf{t}^* 映射到超平面后的向量表示为 \mathbf{h}_\perp^* 和 \mathbf{t}_\perp^* , 如果该三元组成立, 那么 \mathbf{h}_\perp^* 和 \mathbf{t}_\perp^* 应在误差很小的情况下与翻译向量 \mathbf{d}_r^* 联系起来. 据此, 衡量三元组在表示空间中成立可能性的得分函数修订为 $(\mathbf{h}_\perp^* + \mathbf{r}^* - \mathbf{t}_\perp^*)^T \mathbf{D}_r^* (\mathbf{h}_\perp^* + \mathbf{r}^* - \mathbf{t}_\perp^*)$. 实验验证时, 采用了 2 个真实公开数据集 WordNet 和 Freebase, 对链路预测和三元组分类 2 项基准任务开展了全面评测. 实验结果表明, TransAH 在诸多性能指标上均优于其他知识表示方法, 从而验证了所提方法的有效性和优越性.

1 相关工作

知识图谱的提法是在谷歌知识图谱 (Google knowledge graph) 项目中首次披露的. 虽然知识图谱的概念较新, 但它并非一个全新的领域. 早年, Berners-Lee 就提出了数据链接 (linked data) 的思想, 为迎接语义网的到来做好准备.

目前, 知识图谱已经被应用在语义搜索、智能问答等诸多领域, 相对成熟的国外产品包括谷歌公司使用的 Knowledge Vault、苹果公司使用的 Wolfram Alpha 智能计算引擎以及 Freebase 和 YAGO 等, 针对中文产品则包括百度“知心”和搜狗“知立方”等.

当前, 面向知识图谱的表示学习的研究主要集中在基于翻译的模型上, 代表性工作主要包括 TransE^[5], TransH^[6], TransR^[7], CTransR^[7], TransD^[8] 等方法.

基于翻译的模型认为, 对每个三元组 (h, r, t) , 其中的关系 r 是从头实体向量 \mathbf{h}^* 到尾实体向量 \mathbf{t}^* 的一个翻译 (translation) 操作. 据此, Bordes 等人^[5] 率先提出了 TransE 知识表示方法. TransE 通过欧氏距离上的偏移量来衡量计算实体之间的语义相似度, 是一种简单基本的知识表示方法. 它的优化目标是尽量使得 $\mathbf{h}^* + \mathbf{r}^* = \mathbf{t}^*$, 因此相应模型学习的得分函数是 $f_r^*(\mathbf{h}^*, \mathbf{t}^*) = \|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*\|^2$, 其中 $\|\cdot\|^2$ 是 \cdot 的 2 范数, 即欧氏距离.

相较之前的模型, TransE 方法在性能和效果上均取得较好的结果^[9]. 但是, TransE 方法由于模型相对简单, 存在无法区分和处理实体之间一对多、多对一以及多对多等复杂关系的问题. 针对此问题, 后续有 TransH, TransR 和 CTransR 等方法提出. 在相同真实数据集上的实验表明, 这些方法从一定程度上解决了复杂关系的表征问题.

其中, TransH 方法建立了一个面向关系的超平面, 它由一个法向量 \mathbf{n}_r^* 和翻译向量 \mathbf{r}^* 表示, 实体向量 \mathbf{h}^* 和向量 \mathbf{t}^* 首先被投影到关系的超平面, 得到向量 $\mathbf{h}_\perp^* = \mathbf{h}^* - \mathbf{n}_r^{T*} \mathbf{h}^* \mathbf{n}_r^*$ 和 $\mathbf{t}_\perp^* = \mathbf{t}^* - \mathbf{n}_r^{T*} \mathbf{t}^* \mathbf{n}_r^*$. 因而, TransH 的优化目标变为 $\mathbf{h}_\perp^* + \mathbf{r}^* = \mathbf{t}_\perp^*$, 其相应的得分函数修改为 $f_r^*(\mathbf{h}^*, \mathbf{t}^*) = \|\mathbf{h}_\perp^* + \mathbf{r}^* - \mathbf{t}_\perp^*\|^2$.

在 TransE 和 TransH 中, 实体和关系存在于同一个空间中. 然而, 实体和关系本质上是不同的客观

① 自反, 即同一个实体和自身的关系; 一对多, 即一个头实体通过同一个关系对应多个尾实体; 多对一, 即多个头实体通过同一个关系对应一个尾实体; 多对多, 即多个头实体通过同一个关系对应多个尾实体.

事物,所以将它们放置于同一个空间中描述是不恰当的.为此,TransR 和 CTransR 希望通过建立一个映像矩阵 \mathbf{M}_r^* 和一个向量 \mathbf{r}^* 来表示每一个关系 r .具体地,TransR 将实体 \mathbf{h}^* 和 \mathbf{t}^* 通过矩阵映射到关系向量 \mathbf{r}^* 的层次上,得到 $\mathbf{M}_r^* \mathbf{h}^* + \mathbf{r}^* = \mathbf{M}_r^* \mathbf{t}^*$,也即 TransR 的优化目标.

CTransR 是 TransR 的扩展,它将多个头尾实体聚集为一类得到每一类特有的关系向量.它在取得进步的同时,仍有缺点:对于 1 个关系 r ,所有的实体共享一个映射矩阵 \mathbf{M}_r^* .注意到,通过一个关系连接的实体类型和属性是不同的,映射是实体与关系之间的交互过程,所以映射矩阵仅由关系确定是不合理的.此外,在 CTransR 的学习过程中,矩阵与向量的相乘使计算量激增,其参数数量也比 TransE 和 TransR 多.因此,CTransR 由于过高的复杂性而不适用于大规模知识图谱.

新近的研究成果 TransD 方法为每个向量和关系定义了 2 个向量:一个向量用来表示实体或关系的含义;另一个向量表示 1 个实体是如何映射到 1 个关系向量空间,称为映射向量,用以生成映射矩阵.因此,每一个实体关系对都会有一个唯一的映射矩阵.从而,TransD 的优化目标变成 $\mathbf{M}_r^* \mathbf{h}^* + \mathbf{r}^* = \mathbf{M}_r^* \mathbf{t}^*$;另外,TransD 以向量操作取代了之前的矩阵与向量的乘法操作,因而具有较高的计算效率,可应用到大型知识图谱上.

除了基于翻译的模型之外,结构化嵌入模型^[10]是较早的知识表示方法,每个实体用 d 维的向量表示,所有实体被投影到同一个 d 维向量空间中.单层神经网络模型^[11]是对结构化嵌入模型的进一步改进,采用了单层神经网络的非线性操作.语义匹配能量模型^[12-13]提出了更复杂的操作,刻画实体与关系的内在联系.潜变量模型^[14-15]提出利用基于关系的双线性变换,刻画实体和关系之间的二阶联系.神经张量网络(nerual tensor network, NTN)模型^[11]用双线性向量取代传统神经网络中的线性变换层.矩阵分解是得到低维向量表示的重要途径,基于这种模型的代表方法是 RESACL^[16-17].关于这些模型的更加详尽的细节和对照,请感兴趣的读者参考综述文献^[9].

关于知识图谱,其自动化构建也是当前的一个持续性的研究热点.一般认为,知识图谱的自动化构建涉及了信息抽取、知识融合和知识加工 3 个主要阶段,构建过程中涉及的关键技术则包括命名实体识别、关系抽取、实体链接和信息融合等.关

于知识图谱自动化构建技术,感兴趣的读者可以参考综述文献^[18].

2 TransAH 知识表示方法

本节详细介绍了所提改进的知识图谱表示方法 TransAH,解释了其原理与算法. TransAH 同属于基于翻译模型的方法,它针对经典的知识表示方法 TransE 的 2 个缺陷,分别采用了自适应的度量方法和超平面模型来加以解决,并将 2 个想法同时集成在 1 个模型框架下,实现了快速的求解与计算.

2.1 自适应的度量方法

关于缺陷 1,究其产生的本质原因是 TransE (以及其他绝大多数基于翻译的方法)采用了同一个朴素的优化目标,即 $\mathbf{h}^* + \mathbf{r}^* = \mathbf{t}^*$;换句话说,TransE 等方法在实体和关系通过不同的法则进行映射后,均采用同样的得分函数对目标进行优化;它们共有的得分函数如下:

$$f_r^*(\mathbf{h}^*, \mathbf{t}^*) = \|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*\|^2 = (\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*)^T (\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*).$$

从上式可见,该得分函数选用了欧氏距离作为表示空间中的差异度量,在空间形态上表现为一个球形等势面.然而,观察到在实际应用中,每个特征维度对于各种关系的影响程度是不同的,统一考虑所有特征维度的权重将直接影响知识表示的效果.因此,这种距离度量在实践应用中有失灵活,导致构建的模型会包含“不必要”的错误,进而使得知识表示的精度下降.

为此,考虑在得分函数中加入一个权重矩阵,将欧氏距离转换为加权欧氏距离作为表示空间中的差异度量,即,通过权重矩阵来灵活控制和确定各个特征维度的重要程度实现自适应的表示空间中的距离度量.

一种设置方法是使用一般权重矩阵 \mathbf{W}_r^* ^[19],其对应的改进后的得分函数为

$$f_r^*(\mathbf{h}^*, \mathbf{t}^*) = (|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*|)^T \mathbf{W}_r^* (|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*|),$$

其中,

$$|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*| = (|\mathbf{h}_1^* + \mathbf{r}_1^* - \mathbf{t}_1^*|, |\mathbf{h}_2^* + \mathbf{r}_2^* - \mathbf{t}_2^*|, \dots, |\mathbf{h}_k^* + \mathbf{r}_k^* - \mathbf{t}_k^*|),$$

其中, k 是特征维的长度.

然而,注意到一般权重矩阵 \mathbf{W}_r^* 在生成过程中效率十分低下,导致模型训练时间成本较大(参考 3.3 节中的实验结果).此外,注意到,现有工作中的 TransR 和 TransD 方法,采用转换矩阵将头实体和尾实体映射到关系空间中,从而可以更好地对复杂

关系进行区分,提高知识表示精度. 相较而言,直接使用权重矩阵 W_r^* 将实体与关系同时映射到关系空间中,徒增了模型的复杂度,除此之外,将 W_r^* 每轮进行归一化的方法并不会提高算的精确度,反而增加了运行的时间开销.

鉴于此,化繁就简,提出使用对角权重矩阵 D_r^* ,而非一般权重矩阵 W_r^* ,以实现自适应的距离度量. 具体地,定义对角权重矩阵 $D_r^* = \text{diag}(\omega_1, \dots, \omega_i, \dots, \omega_k)$,其中 $\text{diag}()$ 表示 D_r^* 是 1 个对角矩阵,第 i 个特征维的重要性由参数 ω_i 表示, $1 \leq i \leq k$.

相比于经典的 TransE 方法,引入对角权重矩阵的模型具有至少 3 个方面的优势:

1) 对角权重矩阵 D_r^* 可实现将欧氏距离转化为加权欧氏距离,将球形等势面转换为更为灵活的椭圆等势面. 球形等势面中的度量准则只是将越靠近中心的三元组作为正确三元组,容易包含不必要的错误. 举例如图 1 中,又表示正确匹配的尾实体,圆表示错误匹配的尾实体. 图 1(a)中采用球形等势面,出现了 7 个错误;采用椭圆等势面后,可以看到图 1(b)中有 4 个错误被避免了. 也就是说,通过优化对角矩阵 D_r^* 中的变量,就可以对固定的球形等势面转换为有一定伸缩性的椭圆等势面,从而在避开了错误实体的同时包含更多正确实体,提升了知识表示的能力.

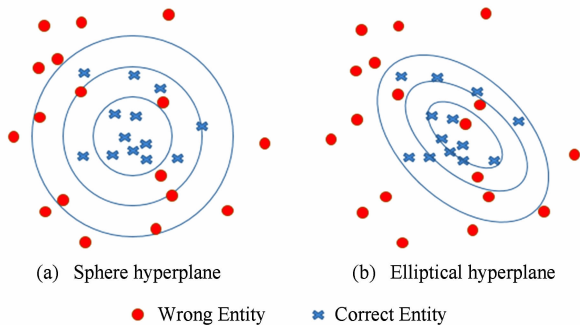
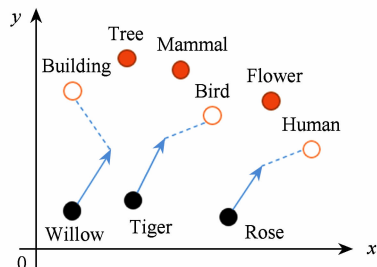


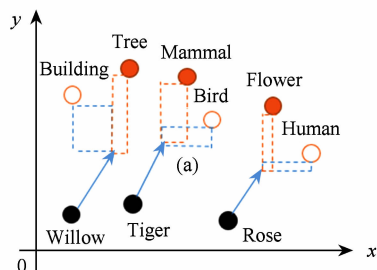
Fig. 1 The comparison on sphere and elliptical hyperplane
图 1 球形与椭圆等势面的比较

2) 一个关系仅由若干个特定的潜在特征维度影响,其他维度会成为干扰. 传统的如 TransE 等基于翻译的方法同等对待各个特征维度,无法滤除无关维度的干扰. 采用对角权重矩阵 D_r^* 则可以认为特征维赋予合理的自适应的权重,即 $\omega_1, \omega_2, \dots, \omega_n$,不同的特征维 i 由 ω_i 来控制其权重. 举例如图 2 所示,实心圆是正确匹配的结果,空心是错误的匹配结果,箭头表示 type_of 关系. 图 2(a)中,由于将某 2 个维度同等对待,导致 3 个头实体都匹配到

了不正确的尾实体,比如实体 Willow 匹配到了实体 Building;图 2(b)中,对这 2 个表示维度赋予权重,尤其是增加了 y 轴的权重,降低了 x 轴的权重,知识表示得以修正和优化,比如实体 Willow 正确匹配到了实体 Tree.



(a) Equally-treated feature dimension



(b) Weigh-assigned feature dimension

○ Wrong Tail Entity ● Correct Tail Entity ● Head Entity

Fig. 2 The comparison on weighting feature dimensions
图 2 特征维度权重化比较

3) 不选用一般权重矩阵 W_r^* 来为特征维赋予权重,主要是因为知识表示中比较不同特征维度的意义是微乎其微的,甚至可能带来负面的影响. 譬如,比较人物实体的“国籍”和“性别”是没有实际意义. 因此,落实到矩阵 W_r^* 中,其变量 ω_{xy} 当 $x \neq y$ 时是没有意义的,认为仅用对角矩阵即可完全表达对特征维的权重化. 同时,只需保证 D_r^* 中的每一个变量非负,无需再像使用 W_r^* 时那样对 $|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*|$ 加上绝对值来防止负的分函数降低整体得分函数的值. 此外,对角权重矩阵 D_r^* 的应用相较于 W_r^* ,大幅度降低了模型的复杂度,因而显著提升了学习计算效率.

为了获得最终的表示模型,其中的对角权重矩阵 D_r^* 需要通过训练不断优化得到. 具体实现中,为保证 D_r^* 的非负性,对其初始值设置为

$$D_r^* = \text{diag} \left(\sum_{(h', r', t') \in \Delta'} (|\mathbf{h}' + \mathbf{r}' - \mathbf{t}'|^2) - \sum_{(h^*, r^*, t^*) \in \Delta} (|\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*|^2) \right).$$

综上,得到基于加权欧氏距离的自适应得分函数:

$$f_r^*(\mathbf{h}^*, \mathbf{t}^*) = (\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*)^T D_r^* (\mathbf{h}^* + \mathbf{r}^* - \mathbf{t}^*).$$

2.2 面向关系的超平面映射模型

考虑缺陷 2, TransE 模型结合自适应度量方法后仍然对处理复杂关系的能力较弱. 其主要原因是因为 TransE 将实体和关系都表示在同一个平面中, 无法理清实体和关系之间的相互影响, 进而造成无法区分复杂关系的问题. 举例如图 3 所示, 对于 2 个具有相同关系的事实知识——布什是美国总统和奥巴马是美国总统, TransE 不能正确区分布什和奥巴马这 2 个实体, 影响了知识表示的准确性.

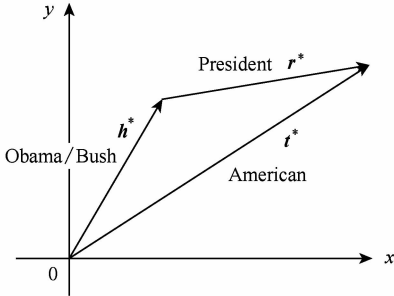


Fig. 3 The effect of TransE method
图 3 TransE 方法效果图

受 TransH^[6]方法的启发, 可通过建立面向关系的超平面模型来改善上述问题. 具体地, 对于一个关系 r , 将一个关系特定的翻译向量 r^* 放在关系特定的超平面 n_r^* (法向量) 上, 而不是与实体表示在同一个空间中. 具体地, 对于一个三元组 (h, r, t) , h^* 和 t^* 首先被映射到超平面 n_r^* 上, 分别表示为 h_\perp^* 和 t_\perp^* ; 在超平面 n_r^* 上, h_\perp^* 和 t_\perp^* 再与翻译向量 r^* 联系起来. 因此, 得分函数进一步修正为 $\|h_\perp^* + r^* - t_\perp^*\|^2$. 通过约束 $\|n_r^*\|_2 = 1$, 容易得到:

$$\begin{cases} h_\perp^* = h^* - n_r^{*T} h^* n_r^*, \\ t_\perp^* = t^* - n_r^{*T} t^* n_r^*. \end{cases}$$

于是, 综合上式, 可以得到应用超平面模型后的得分函数, 即:

$$f_{r^*}(h^*, t^*) = \| (h^* - n_r^{*T} h^* n_r^*) + r^* - (t^* - n_r^{*T} t^* n_r^*) \|^2.$$

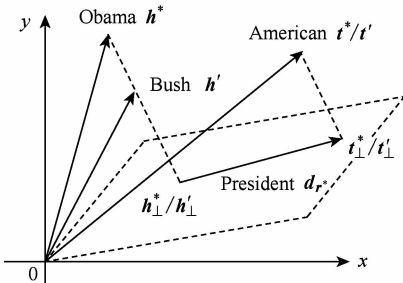


Fig. 4 The effect of hyperspace model
图 4 超平面模型方法效果图

在运用了超平面模型之后, 举例如图 4 所示, 奥巴马和布什这 2 个人物实体通过不同的映射向量投影到“总统”关系的超平面上, 从而得以区分, 保证了知识表示的准确性和可推理性.

2.3 TransAH 的模型与训练

在 2.1 节和 2.2 节内容的基础上, 将自适应度量方法和超平面模型结合起来, 置于一个统一的模型框架下面, 得到了最终得分函数:

$$f_{r^*}(h^*, t^*) = ((h^* - n_r^{*T} h^* n_r^*) + r^* - (t^* - n_r^{*T} t^* n_r^*))^T \cdot$$

$$D_{r^*} \cdot ((h^* - n_r^{*T} h^* n_r^*) + r^* - (t^* - n_r^{*T} t^* n_r^*)).$$

由于本方法同属于基于翻译的模型, 并结合自适应性(adaptive)度量和超平面(hypersphere)的思想, 因此取名为 TransAH, 并将在第 3 节的实验部分与其他当前主流模型及方法进行横向评测.

模型训练中, 采用基于差距的排序误差(margin-based ranking error)函数作为训练模型的优化目标函数:

$$\min \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'} [f_{r^*}(h^*, t^*) + \gamma - f_{r'}(h', t')]_+, \quad (1)$$

其中, $[\cdot]_+ = \max(0, \cdot)$, Δ 是正确的三元组(正三元组^①)的集合, Δ' 是错误的三元组(负三元组)的集合(通过打乱已有黄金三元组得到), γ 是区分正负三元组的差距. 因此, 该优化目标函数的主要目的是将不正确的三元组和正确的三元组进行最大限度的分离.

同时, 在优化式(1)时, 还需要考虑向量中的 3 个约束条件, 主要包括:

$$\forall h, t \in E, \|h^*\|_2 \leq 1, \|t^*\|_2 \leq 1, \quad (2)$$

$$\forall r \in R, |n_r^{*T} r| / \|r^*\|_2 \leq \epsilon, \quad (3)$$

$$\forall r \in R, \|r^*\|_2 = 1, \quad (4)$$

其中, 式(2)为保证实体向量长度不大于 1, 式(3)保证了翻译向量 r^* 在超平面上, 式(4)保证了超平面为单位法向量. 同时, 还要保证对角矩阵 D_{r^*} 满足:

$$\|D_{r^*}\|_F^2 \leq 1.$$

另外, 不直接带着上述这些约束优化式(1)中的损失函数, 而是通过软约束的方法将式(1)转化为不受约束的形式. 因此, 最终得到优化目标函数所示:

$$\min \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'} [f_{r^*}(h^*, t^*) + \gamma - f_{r'}(h', t')]_+ + \lambda \left(\sum_{r \in R} \|D_{r^*}\|_F^2 \right) + \eta \left(\sum_{e \in E} \|e\|^2 + \sum_{r \in R} \left[\frac{(n_r^{*T} r)^2}{\|r^*\|^2} - \epsilon^2 \right] \right),$$

① 又称黄金三元组, 即知识图谱中已经包含的实体关系三元组.

其中,目标函数的第 2 项与第 3 项是对约束条件的表示,用于防止过拟合,而 λ 和 η 是控制软约束的参数。

具体模型训练时,采用了经典的随机梯度下降(stochastic gradient descent)法来优化上述目标函数。正确的三元组集在训练过程中会循环多次,当遇到正确三元组时,同时会随机生成不正确三元组;在一小批次的训练后,梯度以及模型的参数会进行自动更新。

3 实 验

本节介绍 TransAH 知识表示方法的实验验证,主要在 2 个工作上验证和评估了 TransAH 方法——链路预测和三元组分类。这 2 个工作从不同的角度评估了模型预测不可见三元组精确度的能力,对应于不同的应用场景。首先,介绍了这 2 项工作的评价准则,实验实现的具体配置以及相应的实验结果;然后,分析了 TransAH 方法的实验效果,着重考察了 TransAH 方法在表征复杂关系的能力和训练效率,并与其他知识表示方法进行比较。

3.1 实验准备

1) 抽样策略。进行训练时,需要基于黄金三元组构建负三元组。之前的方法只是通过随机打乱黄金三元组来获得负的三元组,例如,在 TransE 中,对于一个黄金三元组 (h, r, t) ,那么一个负的三元组 (h', r, t') 是由从实体集 E 中随机抽取一对实体 (h', t') 。但是这种方法对于多元关系,会存在将原本正确的三元组标记为错误的三元组的情况,因此,改用伯努利抽样的策略。

在打乱一个三元组时,根据关系的映射关系为代替头和尾实体设置了不同的概率,比如说一对一、多对一和多对多等。一般倾向于给一对多的关系中更高的概率代替头实体,在多对一关系中更高的概率代替尾实体。用这种方法,产生错误负三元组的概率大大降低了。

具体地,一个关系 r 的所有三元组中,首先可以获得以下 2 组数据:1) 每一个头实体对应的尾实体的平均数量,记为 tph (# tail entities per head entity);2) 每一个尾实体对应的头实体的平均数量,记为 hpt (# head entities per tail entity)。然后,用参数

$$p = \frac{tph}{tph + hpt}$$

的伯努利分布来取样。对于一个关系 r 给定的黄金三元组 (h^*, r^*, t^*) ,在进行打乱时,以概率 p 代替它的头实体,以概率 $1-p$ 代替它的尾实体。

同样,亦可得到区分出关系类型的方法。对于每一个关系 r ,计算了每一个头实体的尾实体的平均数量 tph_r , 每一个尾实体的头实体的平均数量 hpt_r 。具体地,如果 $tph_r < 1.5$ 且 $hpt_r < 1.5$,那么关系 r 是一对一的;如果 $tph_r \geq 1.5$ 且 $hpt_r \geq 1.5$,那么关系 r 是多对多的;如果 $hpt_r < 1.5$ 且 $tph_r \geq 1.5$,那么关系 r 是一对多的;如果 $hpt_r \geq 1.5$ 且 $tph_r < 1.5$,那么关系 r 是多对一的。

2) 平均三元组的数量。平均三元组的数量 $ATPE$ (average triple number per entity) 衡量了数据多样性和复杂性。直观地看,数量越多的三元组导致了越复杂和越稠密的知识图谱结构。为了表达更加复杂的图谱结构,实体的分配也会更加多样化和复杂化。因此,总体上说,知识表示效果会随着 $ATPE$ 的增大而逐步退化,毕竟越高的 $ATPE$ 代表意味着更为多样和复杂的实体关系情形。

3.2 链路预测

链路预测的主要任务是,对于一个丢失了头实体或尾实体的三元组 (h, r, t) ,给定 (h, r) 预测 t ,或者给定 (r, t) 预测 h 。这项测试任务着重于对知识图谱中的候选实体集进行排序,而不是直接获得一个最好的答案。

本组实验选用了 TransE 使用的 2 个数据集: WordNet 中的子集 WN18 和 Freebase 中的一个相对稠密的子集 FB15K,其中的实体也都包括在维基数据库中。关于实验数据集的基本统计信息,如表 1 所示:

Table 1 Statistics of Datasets

表 1 数据集统计信息

Statistics	WN18	FB15K	WN11	FB13
# Rel	18	1 345	11	13
# Ent	40 943	14 951	38 696	75 043
# Train	141 442	483 142	112 581	316 232
# Valid	5 000	50 000	2 609	5 908
# Test	5 000	59 071	10 544	23 733
ATPE	3.70	39.61	3.25	4.61

1) 评价准则。为了更好地与 TransE 等知识表示模型进行对比,采用和 TransE 相同的评价准则。对于每一个测试三元组 (h, r, t) ,将尾实体 t 用每一个知识图谱中的实体 e 来代替,同时用得分函数

$f_r^*(h, e)$ 计算损坏的三元组 (h, r, e) 的相应得分. 用升序对这些分数排序, 得到了正确三元组的排序得分. 同样, 也能得到打乱头实体 h 的三元组 (h, r, t) 的排序得分.

将所有的测试三元组进行综合, 有 2 个度量准则: ①黄金三元组的平均排序得分, 记为 *MeanRank*; ②黄金三元组排序不大于 10 的比例, 记为 *Hits@10*. 注意到, 如果一个损坏的三元组在知识图谱中存在, 即该三元组实际上是正确的, 那么将它排在原始三元组之前也是合理的. 为了消除这个因素的影响, 实验中在得到每一个测试三元组的排名得分之前, 将上述产生“干扰”的损坏三元组从训练集、验证集和测试集中去除, 从而保证了该损坏的三元组不属于任何数据集. 这个设置称为“Filt”设置, 而未经上述处理的实验设置成为“Raw”设置. 在这 2 种实验设置中, 一个更低的 *MeanRank* 和更高的 *Hits@10* 意味着更好的实验效果.

2) 实验实现. 训练 TransAH 时, 在随机梯度下降法中使用的学习速率 $\alpha = \{0.002, 0.005, 0.01\}$, 差距 $\gamma = \{0.25, 0.5, 1, 2\}$, 表示维度 $k = \{50, 75, 100\}$, 权重 $\eta = \{0.05, 0.0625, 0.25, 1.0\}$, 训练批量的大小 $B = \{20, 75, 200, 1,200, 4,800\}$. 最优参数由验证集决定. 用“unif.”表示传统的等概率替代头实体或者尾实体的方法, 用“bern.”来表示使用伯努

利抽样策略的方法, 即用不同的概率分别来代替头实体和尾实体.

在等概率抽样中, 训练 TransAH 最优的参数配置如下: 在 WN18 数据集上, $\alpha = 0.01, k = 50, \gamma = 1.0, \eta = 0.25, B = 75$; 在 FB15K 数据集上, $\alpha = 0.005, k = 50, \gamma = 0.5, \eta = 0.05, B = 1200$. 在伯努利抽样中, 训练 TransAH 最优的参数配置如下: 在 WN18 数据集上, $\alpha = 0.01, k = 50, \gamma = 1.0, \eta = 0.25, B = 1200$; 在 FB15K 数据集上, $\alpha = 0.005, k = 100, \gamma = 0.25, \eta = 1.0, B = 4800$. 对于每一个数据集, 本实验将所有训练三元组迭代了 500 次.

3) 实验结果. 实验结果在表 2~4 中列出, 每组实验的最优值进行了加粗以突出显示.

如表 2 所示, 在 WN18 中, TransE 和 TransAH 等方法甚至是原始的, 对关系没有进行翻译操作的 TransE 方法, 在 *MeanRank* 这个度量上, 都比其他方法要好. 这可能是因为 WN18 中关系的数量比较少, 所以忽略掉不同类型的关系也是合理的. 但在 FB15K 中, TransAH 方法比所有其他方法表现都要好. FB15K 是一个多样复杂的实体关系图谱, 它的 *ATPE* 值也是这些数据集中是最高的. 在这个数据集中, TransAH 的实验表现是最好的, 说明 TransAH 方法在多样复杂的实体关系表示方面更具优势.

Table 2 The Average Results of Link Prediction

表 2 链路预测的平均预测结果

Method	WN18				FB15K			
	<i>MeanRank</i>		<i>Hits@10</i>		<i>MeanRank</i>		<i>Hits@10</i>	
	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt
Unstructured ^[13]	315	304	35.3	38.2	1074	979	4.5	6.3
RESCAL ^[17]	1180	1163	37.2	52.8	828	683	28.4	44.1
SE ^[10]	1011	985	68.5	80.5	273	162	28.8	39.8
SME (Linear) ^[12]	545	533	65.1	74.1	274	154	30.7	40.8
SME (Bilinear) ^[12]	526	509	54.7	61.3	284	158	31.3	41.3
LFM ^[15]	469	456	71.4	81.6	283	164	26.0	33.1
TransE ^[5]	263	251	75.4	89.2	243	125	34.9	47.1
TransH ^[6]	401	388	73.0	82.3	212	87	45.7	64.4
TransA ^[19]	405	392	82.3	94.3	155	74	56.1	80.4
TransR ^[7]	238	225	79.8	92.0	198	77	48.2	68.7
TransAH (unif.)	181	170	71.5	81.7	150	68	55.2	81.2
TransAH (bern.)	182	171	72.8	82.9	141	65	58.1	82.4

Table 3 Hits@10 of Each Type of Relations in FB15K

表 3 关于 FB15K 各类关系的 Hits@10 值

Method	Predicting Left (Hits@10)				Predicting Right (Hits@10)			
	1-1	1-n	n-1	n-n	1-1	1-n	n-1	n-n
Unstructured ^[13]	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE ^[10]	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME ^[12] (bilinear)	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE ^[5]	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH ^[6]	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransR ^[7]	78.8	89.2	34.1	69.2	79.2	37.4	90.4	72.1
TransAH (unif.)	82.9	94.0	36.1	72.5	82.5	44.1	93.9	75.5
TransAH (bern.)	82.7	94.0	36.4	72.2	82.5	44.2	93.8	75.3

Table 4 Hits@10 of Several Relations in TransE and TransAH

表 4 TransE 和 TransAH 关于若干关系的 Hits@10 值

Relation	Hits@10 (TransE/TransAH)	
	Head	Tail
football_position/players	100/ 100	16.7/ 66.7
production_company/films	65.6/ 95.7	9.3/ 41.7
director/film	75.8/ 93.4	50.5/ 90.6
disease/treatments	33.3/ 66.6	100/ 100
person/place_of_birth	30.0/ 54.5	72.1/ 90.3
film/production_companies	11.3/ 38.1	77.6/ 96.1
field_of_study/students_majoring	24.5/ 67.9	28.3/ 71.7
award_winner/awards_won	40.2/ 95.4	42.8/ 94.1
sports_position/players	28.6/ 100	64.3/ 91.4
person/sibling_s	21.1/ 68.4	21.1/ 73.7
person/spouse_s	18.5/ 57.4	18.5/ 55.6

与 TransE 相比, TransAH 在 WN18 上改进了 31.2%, 而在 FB15K 上进步了 36.2%。这个比较进一步显示了 TransAH 相比 TransE 在表达多样复杂关系的场合应用中有更大的优势。同时, 与 TransA 比较, TransAH 同样取得了一定的进步。这些改进主要是由于 TransAH 采用了相较于经典方法更加灵活的度量标准, 并应用了超平面模型, 因此能够更好地表达自反、一对多、多对一和多对多等复杂关系, 进而更好地支持准确推理。

为了进一步证实这一观点, 深入挖掘分析了不同关系不同映射类型的相应结果, 具体数值如表 3 和表 4 中所示。在 1345 个关系中, 24% 的关系是一对一的, 23% 的关系是一对多的, 29% 的关系是多对一的, 24% 的关系是多对多的。表 3 中, “Predicting Left” 和 “Predicting Right” 分别表示由关系和尾实体预测头实体以及由关系和头实体预测尾实体。相

较于 TransE 方法, TransAH 在一对多、多对一、多对多关系表达上有显著的改善。尤其出乎预想的是, 在一对一关系中, TransAH 同样有巨大的进步(超过 80%)。这可能受益于知识图谱本身的某些特征, 即实体与关系相联系, 某一部分表示的更好使得整体结果得到较大改进。表 4 中, 自上而下显示了在一对多、多对一、多对多和自反等几个典型复杂关系上的 Hits@10 值。与 TransE 方法相比, TransAH 在这些关系中的改善同样可圈可点, 尤其是在 person/sibling_s 和 person/spouse_s 关系上, TransAH 在 Hits@10 上的提升超过 56.4%。

3.3 三元组分类

三元组分类是确定一个给定三元组 (h, r, t) 是否正确, 其主要任务是对一个三元组进行“正确”和“错误”的二元分类。

实验中, 首先使用了 WordNet 的子集 WN11 和 Fresbase 中的子集 FB13; 由于 WN11 和 FB13 包含的关系数过少, 还额外使用了包含更多关系的 FB15K 数据集。关于实验数据集的基本统计信息, 详见表 1。

1) 评价准则。本组实验遵循了 NTN 模型所采用的评估准则。进行分类时, 需要生成负的标签, 通过将黄金三元组打乱得到负的三元组, 具体方法同样可参照 3.1 节中的抽样策略。

分类的决定规则是, 对于一个三元组 (h, r, t) , 如果其得分函数比一个关于关系 r 的给定阈值 σ_r 低, 那么预测为正确, 反之为错误。特定关系的阈值 σ_r 由验证集得到最大分类精度时的阈值决定。

此外, 为了证实使用对角权重矩阵 D_r^* 比使用一般权重矩阵 W_r^* 在计算效率方面的提升, 将包含权重矩阵 W_r^* 的方法标记为 TransAH-W, 一并

进行了实验,以比较实验的运行时间和实验结果.

2) 实验实现. TransAH 设置的学习速率 $\alpha = \{0.001, 0.005, 0.01, 0.1\}$, 差距 $\gamma = \{1.0, 2.0\}$, 表示维度 $k = \{20, 50, 100\}$, 训练批量的大小 $B = \{30, 120, 480, 1920\}$.

TransAH 在伯努利抽样的实验中,最优的参数配置为:在 WN11 中, $\alpha = 0.01, k = 20, \gamma = 2.0, B = 120$; 在 FB13 中, $\alpha = 0.001, k = 100, \gamma = 2.0, B = 30$; 在 FB15K 中, $\alpha = 0.005, k = 100, \gamma = 2.0, B = 480$. TransAH 在等概率抽样的实验中,最优的参数配置为:WN11 中, $\alpha = 0.01, k = 100, \gamma = 2.0, \eta = 0.25, B = 4800$; FB13 中, $\alpha = 0.001, k = 100, \gamma = 0.25, \eta = 0.0625, B = 4800$; FB15K 中, $\alpha = 0.01, k = 100, \gamma = 0.25, \eta = 0.0625, B = 4800$. 循环次数限制在 500 次.

3) 实验结果. 分类精度结果如表 5 所示. 在 WN11 和 FB13 中, TransAH 比其他方法出色; 而在 FB13 中, NTN 模型表现同样出色. 但是在更大的数据集 FB15K 中, TransE 和 TransAH 则都比 NTN 出色. 这说明基于翻译的模型相对更适用于大规模的知识图谱. 注意到 FB15K 中关系的数量 (1345) 远远大于 FB13 中关系的数量 (13), 而实体的数量却十分接近, 说明 FB13 是一个密度很大的子数据集, 实体间存在很强的联系. 在这个数据集中, NTN 通过张量和非线性转换方法来对复杂的联系进行建模有一定优势. 但是, 在 FB15K 这个稀疏的子图谱中, 采用灵活度量法和应用超平面模型十分适用, 而 NTN 却并不适用. 不仅如此, 考虑到运行时间, NTN 的消耗时间比 TransE 和 TransAH 都要高. 另外, 在所有 3 个数据集中, 使用伯努利抽样也有一定的作用.

Table 5 Accuracy and Running Time of Triplet Classification of Different Models

表 5 不同模型的三元组分类精度及运行时间

Method	WN11	FB13	FB15K
Distant ^[10]	53.0	75.2	
Hadamard ^[17]	70.0	63.7	
SLM ^[12]	69.9	85.3	
SME ^[12] (bilinear)	73.8	84.3	
NTN ^[11]	70.4	87.1	66.5 (40 h)
TransE ^[5]	75.87	81.5	79.7
TransH ^[6]	78.80	83.3	87.7
TransAH (unif.)	85.68	85.51	91.9 (30 min)
TransAH (bern.)	85.2	88.1	92.0 (30 min)
TransAH-W (bern.)	83.8	84.5	89.2 (10 h)

图 5 和图 6 展示了不同类型关系的预测精度. 在不同的数据集中, 不同关系有不同的预测精度, 我们尤其关注精确度较低的关系. 在 WN11 数据集, similar_to 关系的分类精确度仅 75%. 直觉上来讲, similar_to 可以从其他信息中推导出来; 然而, 通过 similar_to 相联系的实体对只有 1672 个, 只占了所有数据的 1.5%, 而预测这个关系需要与实体相关更多信息. 因此, 预测准确度低的主要原因是信息不充分. 在 FB13 中, cause_of_death 和 gender 这 2 个关系的精确度要比其他关系低, 这主要是因为很难从其他信息中预测它们, 尤其是 cause_of_death. 关系 gender 可能从一个人的姓名中推断, 但是我们学习的是每一个名字的向量, 而不是名字中包括的单词, 这导致姓名信息显得没有用处. 综上所述, 通过知识图谱来推导新事实的能力有一定的局限性, 而从纯文本中抽取事实或是一个互补方法.

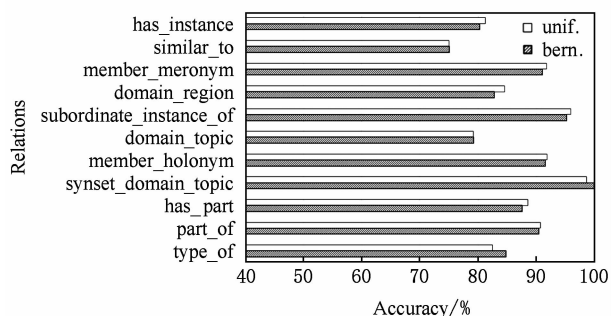


Fig. 5 Prediction accuracy of different relations in WN11

图 5 WN11 中各类关系的预测精度

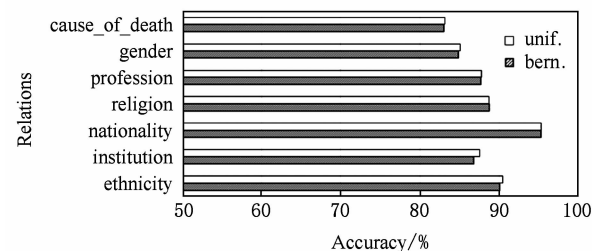


Fig. 6 Prediction accuracy of different relations in FB13

图 6 FB13 中各类关系的预测精度

比较 TransAH 和 TransAH-W 的实验时间和实验效果, 可见 TransAH 的模型训练时间相较 TransAH-W 大大缩短 (从 10h 缩短到 30 min), 但预测精度却反而稍稍占优 (从 89.2% 提升到 92.0%). 这组实验很好地说明了采用对角权重矩阵的有效性和高效性.

为了验证将特征维权重化后的实验效果, 引入“权重差异”的概念. 由于部分维度的权重过小, 可能

存在除法无法计算的情况,因此采用中位数来代替相对较小的权重.于是,权重差异为

$$\frac{\text{MaximalWeight} - \text{MedianWeight}}{\text{MedianWeight}}$$

如图7和图8所示,在WN11和FB13这2个数据集中,各个典型关系的预测精度与权重差异的变化趋势基本是一致的,其中 x 轴罗列了关系的名称,左 y 轴显示了关系的精度(%),右 y 轴是权重差异值.权重差异越大,说明特征维度权重的表达越有意义.可以看到,权重差异依据关系不同会有波动;比如在FB13上,所选典型关系的权重差异多数达到了6以上,但也有部分关系的权重差异在3~4之间.另外,当权重差异增大时,预测精度也相对越大.这说明特征维度权重的应用确实对分类精度的提升起到了促进作用,从而亦证明了为特征维赋予权重的必要性和有效性.

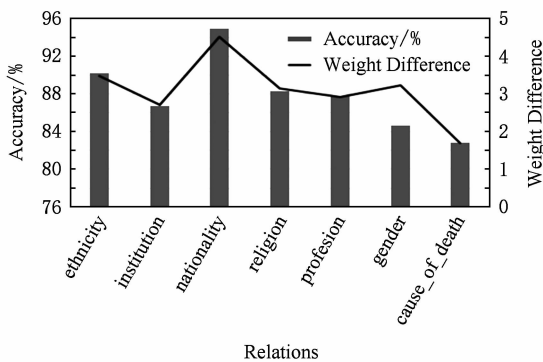


Fig. 7 Accuracy and weight difference of different relations in WN11

图7 WN11中各个关系的精度和权重差异

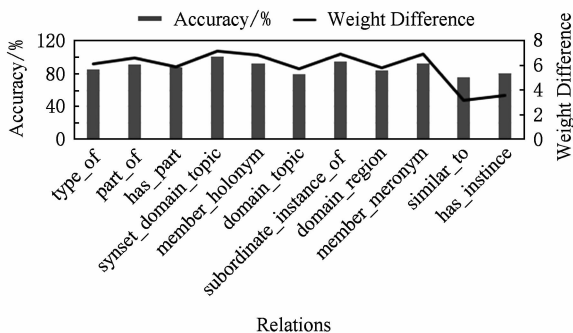


Fig. 8 Accuracy and weight difference of different relations in FB13

图8 FB13中各个关系的精度和权重差异

4 结论和进一步工作

本文主要提出了一种面向知识图谱的改进的知

识表示方法 TransAH. 经典的基于翻译的知识表示方法 TransE 存在距离度量不够灵活和无法处理复杂关系等缺陷,导致知识表示的性能亟待提高.针对第1个缺陷,TransAH采用了一种自适应的度量方法,加入了对角权重矩阵 D ,将得分函数的度量由欧氏距离转换为加权欧氏距离,并实现了为每一个特征维赋予不同的权重.针对第2个缺陷,TransAH应用了面向关系的超平面投影的思想,将头尾实体映射至给定关系的超平面来加以区分.为验证方法的有效性,在基于 WordNet 和 Freebase 的大规模真实数据集上对链路预测和三元组分类这2项任务进行了综合评测.横向比较实验结果表明,TransAH取得最优化的性能,可以应用到真实大规模知识图谱的完善和推理应用中.

下一步,计划对所提 TransAH 方法进行进一步改进,寻求额外的性能提升.注意到 TransD^[8]在对复杂关系进行表示时效果稍优于 TransH,可考虑将其与本文方法相结合,可能会产生更好的结果;但是同时也要考虑到训练实验的计算效率问题,保证面向大规模知识图谱的可扩展性.另外,除了链路预测和三元组分类预测等基础任务,还将致力于考察所提知识表示方法在文本关系知识的抽取、语义实体解析以及基于链路的实体聚类等方面的任务和应用.

参 考 文 献

- [1] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge [C] //Proc of KDD 2008. New York: ACM, 2008: 1247-1250
- [2] Miller G A. WordNet: A lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41
- [3] Wang Haofen. Technology of large scale knowledge graph [J]. Communications of the CCF, 2014, 10(3): 64-68 (in Chinese)
(王昊奋. 大规模知识图谱技术[J]. 中国计算机学会通讯, 2014, 10(3): 64-68)
- [4] Yu Ge, Gu Yu, Bao Yubin, et al. Large scale graph data processing on cloud computing environments [J]. Chinese Journal of Computers, 2011, 34(10): 1753-1767 (in Chinese)
(于戈, 谷峪, 鲍玉斌, 等. 云计算环境下的大规模图数据处理技术[J]. 计算机学报, 2011, 34(10): 1753-1767)
- [5] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data [C] //Proc of NIPS 2013. Cambridge, MA: MIT Press, 2013: 2787-2795

- [6] Wang Zhen, Zhang Jianwen, Feng Jianlin, et al. Knowledge graph embedding by translating on hyperplanes [C] //Proc of AAAI 2014. Menlo Park, CA: AAAI, 2014: 1112-1119
- [7] Lin Yankai, Liu Zhiyuan, Sun Maosong, et al. Learning entity and relation embeddings for knowledge graph completion [C] //Proc of AAAI 2015. Menlo Park, CA: AAAI, 2015
- [8] Ji Guoliang, He Shizhu, Xu Liheng, et al. Knowledge graph embedding via dynamic mapping matrix [C] //Proc of ACL 2015. Stroudsburg PA: ACL, 2015: 687-696
- [9] Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Knowledge representation learning: A review [J]. Journal of Computer Research and Development, 2016, 53(2): 247-261 (in Chinese)
(刘致远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016, 53(2): 247-261)
- [10] Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases [C] //Proc of AAAI 2011. Menlo Park, CA: AAAI, 2011: 301-306
- [11] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion [C] //Proc of NIPS 2013. Cambridge, MA: MIT Press, 2013: 926-934
- [12] Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data [J]. Machine Learning, 2014, 94(2): 233-259
- [13] Bordes A, Glorot X, Weston J, et al. Joint learning of words and meaning representations for open-text semantic parsing [C] //Proc of AISTATS 2012. Cadiz, Spain: JMLR, 2012: 127-135
- [14] Sutskever I, Tenenbaum J B, Salakhutdinov R. Modelling relational data using Bayesian clustered tensor factorization [C] //Proc of NIPS 2009. Cambridge, MA: MIT Press, 2009: 1821-1828
- [15] Jenatton R, Roux N L, Bordes A, et al. A latent factor model for highly multi-relational data [C] //Proc of NIPS 2012. Cambridge, MA: MIT Press, 2012: 3167-3175
- [16] Nickel M, Tresp V, Kriegel H. A three-way model for collective learning on multi-relational data [C] //Proc of ICML 2011. New York: ACM, 2011: 809-816
- [17] Nickel M, Tresp V, Kriegel H. Factorizing YAGO: Scalable machine learning for linked data [C] //Proc of WWW 2012. New York: ACM, 2012: 271-280

- [18] Liu Qiao, Li Yang, Duan Hong, et al. Knowledge graph construction techniques [J]. Journal of Computer Research and Development, 2016, 53(3): 582-600 (in Chinese)
(刘峭, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600)
- [19] Xiao Han, Huang Minlie, Hao Yu, et al. TransA: An adaptive approach for knowledge graph embedding [J/OL]. Computer Science, 2015[2016-09-27]. http://xueshu.baidu.com/s?wd=paperuri%3A%288d5d5c7bc0adb8e1cd42ab11d7e2b31b%29&filter=sc_long_sign&tn=SE_xueshuource_2kduw22v&sc_vurl=http%3A%2F%2Ffarxiv.org%2Fabs%2F1512.01370&ie=utf-8&sc_us=16885378228616607912



Fang Yang, born in 1993. Master candidate. Member of CCF. His main research interests include knowledge graph and big data management, etc.



Zhao Xiang, born in 1986. PhD, lecturer. His main research interests include graph data management and mining, intelligence analytics, etc.



Tan Zhen, born in 1991. PhD candidate. His main research interests include knowledge representation and knowledge graph construction, etc.



Yang Shiyu, born in 1986. PhD. His main research interests include spatial-temporal data management and mining.



Xiao Weidong, born in 1968. PhD, professor, PhD supervisor. His main research interests include intelligence analytics, big data and social computing, etc.